Evaluation of Image Similarity by Histogram Intersection

S. M. Lee,[1] J. H. Xin,[1] S. Westland[2]*

[1]Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[2]School of Design, University of Leeds, Leeds, United Kingdom

Received 1 March 2004; revised 16 August 2004; accepted 16 January 2005

Abstract: Colour is the most widely used attribute in image retrieval and object recognition. A technique known as histogram intersection has been widely studied and is considered to be effective for color-image indexing. The key issue of this algorithm is the selection of an appropriate color space and optimal quantization of the selected color space. The goal of this article is to measure the model performance in predicting human judgment in similarity measurement for various images, to explore the capability of the model with a wide set of color spaces, and to find the optimal quantization of the selected color spaces. Six color spaces and twelve quantization levels are involved in evaluating the performance of histogram intersection. The categorical judgment and rank order experiments were conducted to measure image similarity. The CIELAB color space was found to perform at least as good as or better than the other color spaces tested, and the ability to predict image similarity increased with the number of bins used in the histograms, for up to 512 bins (8 per channel). With more than 512 bins, further improvement was negligible for the image datasets used in this study. © 2005 Wiley Periodicals, Inc. Col Res Appl, 30, 265–274, 2005; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/col.20122

Key words: color imaging; histograms

INTRODUCTION

It is known that color, spatial, and temporal properties are important and need to be considered if the perception of images is to be understood and modeled. For example, correlations between colors over space produce the percept of structures, regions, and shapes and these are not predictable from a consideration of the color of each individual pixel without reference to the spatial relationships of the pixels. Some of the problems of analysing color and spatial properties of images in isolation have been discussed by Caelli and Reye.[1] Wuerger *et al*. have carried out psychophysical experiments, which demonstrate that chromatic and spatial information in images may not be processed independently for certain visual tasks.[2] It is clear, therefore, that a general purpose image-similarity measure should consider at least the spatial and color attributes of the images.

However, color indexing (which ignores the spatial relationships between pixels in an image) is still at the heart of many image-search database systems.[3,4] Histogram-based techniques have been widely studied and are considered to be effective for color-image indexing.[5] A possible explanation for the success of such methods is that color is a critical perceptual cue, and the frequency distribution of colors in an image is robust to confounding factors such as image rotation and even perspective distortion.[5] Histograms are invariant to translation and rotation around the viewing axis and vary slowly with changes of view angle, scale, and occlusion.[6] These features provide robustness with respect to geometric change of the object in the image.[7] It has not been clearly established whether the choice of color space affects the usefulness of the histograms. The purpose of this article is to explore the effect of color space and degree of quantization (the number of histogram bins) on the histogram intersection technique's ability to compute perceptually meaningful image-similarity metrics.

Mehtre *et al*.[8] define the problem of color-image indexing and retrieval in the following statement: "Assume that there are a large number of color images in the database. Given a query image, we would like to obtain a list of images from the database, which are 'most' similar in color to the query image." In defining an effective color-image retrieval strategy, two aspects must be considered: (1) the features (index terms) representing the color information, and (2) the

*Correspondence to: Stephen Westland (e-mail: s.westland@leeds.ac.uk)

Contract grant sponsor: Hong Kong Polytechnic University.

method for measuring the similarity between the features of two images.

One way to develop and test the design of image-similarity models is to measure psychophysical performance for the image-similarity task. Such psychophysical experiments have been carried out and are described in this article. The outputs from these experiments have been used to evaluate the effect of color space and quantization level on model performance. In the following sections, some basic properties of histogram intersection are described, including the nature of some color spaces and the property of quantization.

Histogram Intersection

Given a color space defined by a number (usually three) of axes, the color histogram is obtained by discretizing the image colors and counting the frequency of each discrete color that occurs in the image. Thus, the colors in the image are mapped into a discrete color space containing n colors. A color histogram of image I is an n-dimensional vector, $H_j(I)$, where each element represents the frequency of color j in image I.

Colour indexing recognizes images or image components based upon histogram distributions of the color of pixels, and Swain and Ballard introduced a histogram matching method called Histogram Intersection.[5,6,9] Given a pair of histograms, $H(I)$ and $H(I')$, of images I and I' respectively, each containing n bins, they defined the histogram intersection of the normalized histogram as follows:

$$H(I) \cap H(I') = \sum_{j=1}^{n} \min(H_j(I), H_j(I')) \tag{1}$$

For two images, the larger the value of the histogram intersection, the more similar the image pair is deemed to be. The key issue for histogram-based algorithms is whether the selection of the color space and the quantization level affects the performance of the algorithm.

Colour Spaces

A color space may be defined as a model representing color in terms of a number of intensity values.[10] Six color spaces have been implemented in this study. The CIE 1931 XYZ space is widely used for color measurement and was adopted by the Commission Internationale de l'Éclarage (CIE). However, the RGB system is widely used in machine vision and in many digital-imaging devices. One limitation of most RGB systems is that the values of red (R), green (G), and blue (B) obtained for a given color depend upon the nature of the primaries in a certain imaging device, whereas the XYZ system is said to be a device-independent color space. Human vision is characterized by the ability to detect the perceptual quantities, such as hue, saturation, and intensity (or brightness). Thus, a perceptually relevant color space describes color corresponding to perceived changes in the hue and saturation of a color.[11] Therefore, in addition to XYZ and RGB color spaces, four perceptually inspired color models were used in this study: an opponent space, HSV, CIELAB, and CIELCh color spaces. In the following section, some further details are provided about these color spaces and how they have been used in this study.

RGB color space is an additive system represented in three primaries, R, G, and B. This color space is specified by the chromaticities of its primaries and its white point. The RGB system provides a fast and simple computation, but it is neither a perceptually uniform nor an intuitive color space. Though most RGB values are device dependent, they can be transformed to and from a device-independent color space such as XYZ.[12] The images used in the experiments in this study were obtained in an arbitrary (unknown) RGB format, but were then displayed on a Sony Multiscan CPD-G500 Trinitron monitor that was calibrated to display consistent colors for the same RGB inputs. The particular RGB space that we used was, therefore, that of our Sony monitor.

The RGB values were converted into the hue (H), saturation (S), and value (V) coordinates of the HSV color space[13] using Eq. (2). Thus,

$$\begin{gathered} H = 60(G - B)/(\max(R,G,B) - \min(R,G,B)) \\ \text{if } R = \max(R,G,B) \\ H = 60(2 + B - R)/(\max(R,G,B) - \min(R,G,B)) \\ \text{if } G = \max(R,G,B) \\ H = 60(4 + R + G)/(\max(R,G,B) - \min(R,G,B)) \\ \text{if } B = \max(R,G,B) \\ V = \max(R,G,B) \\ S = 1 - \min(R,G,B)/\max(R,G,B) \end{gathered} \tag{2}$$

so that (H,S,V) is the corresponding point of (R,G,B) in the HSV color space. For $R,G,B \in [0\cdots1]$, the conversion gives $H \in [0\cdots60]$ and $S,V \in [0\cdots1]$.

The monitor that was used for displaying the images was characterized using the GOG model,[14–16] which allows the XYZ values of the colors that were displayed (and defined by RGB values in the space of our monitor) to be computed. The details of the monitor characterization are described in Appendix 1.

The quantities L^*,a^*,b^* of the CIELAB color space were obtained from the tristimulus values, according to standard methods[17] where the neutral point was set to be the white point of the monitor. The polar coordinates C^*_{ab} and h_{ab} were computed from the a^* and b^* values and together with the L^* value comprised the CIELCh space.

To investigate an opponent color space, the XYZ tristimulus values were converted into Smith–Pokorny cone responses,[18] using the linear transform as shown in Eq. (3), and then, the Boynton–MacLeod opponent values were computed using Eq. (4). Thus,

$$\begin{aligned} L &= 0.15515X + 0.54312Y - 0.03286Z \\ M &= -0.15514X + 0.45684Y + 0.03280Z \\ S &= 0.00000X + 0.00000Y + 0.00801Z \end{aligned} \tag{3}$$

TABLE I. The quantization levels and the corresponding number of bins.

	Quantization levels		
	Achromatic axis	Chromatic axes	Total number of bins
Quantize equally	4	4×4	64
	8	8×8	512
	12	12×12	1728
	16	16×16	4096
	20	20×20	8000
Quantize differently	1	8×8	64
	2	16×16	512

and then

$$\begin{aligned} Lum &= L + M \\ RG &= L/(L + M) \\ YB &= S/(L + M) \end{aligned} \tag{4}$$

Equations (3) and (4) allowed the derivation of a light-dark (*Lum*) response and two opponent-color components (*RG* and *YB*).

Quantization

The histogram dimension (the number of histogram bins, n) is determined by the degree of quantization for a given color space. All of the color spaces considered in this study represent color as a three-dimensional vector of real values. A color space of three axes may be quantized into i, j, and k bins for each axis. The histogram can then be represented as an n-dimensional vector whose length is given by the product, $n = ijk$.

To fulfill the scope of this study in comparing different color spaces, the range of each color space was scaled to [0–255, 0–255, 0–255] for the three axes. Thus, for example, if the maximum and minimum L^* values in the dataset of images used in this study were $L_{\max}$ and $L_{\min}$, respectively, then all the L^* values were transformed using the normalization $L^* = 255(L^* - L_{\min})/(L_{\max} - L_{\min})$. Similar normalizations were applied to the dimensions of all of the color space so that each could be treated identically by the subsequent histogram algorithms.

FIG. 1. Carpet database.

FIG. 2. Cork database.

EXPERIMENTAL PROCEDURES

Computational Models

A total of 36 models were implemented and tested, as described in the following sections. Thirty models were tested, where an equal number of quantization levels was used in each of the three axes of each color space. Six color spaces (RGB, XYZ, CIELAB, CIELCh, opponent, and HSV) were investigated, each of them using five quantization models ($4 \times 4 \times 4$, $8 \times 8 \times 8$, $12 \times 12 \times 12$, $16 \times 16 \times 16$, and $20 \times 20 \times 20$).

Swain and Ballard used half the quantization levels for the achromatic component than those used for the chromatic components.[6] Swain and Ballard chose opponent color space specially to allow the intensity axis to be coarsely quantized. To expedite the further investigation of this hypothesis and to compare the performance with the models of equally-quantized in three axes, one representation of the histogram was formed by coarse quantization level in achromatic axis and finer quantization level in the chromatic axes, to generate the same number of levels as the equally quantized histogram. Therefore, the histograms of 64 bins

FIG. 3. Linoleum database.

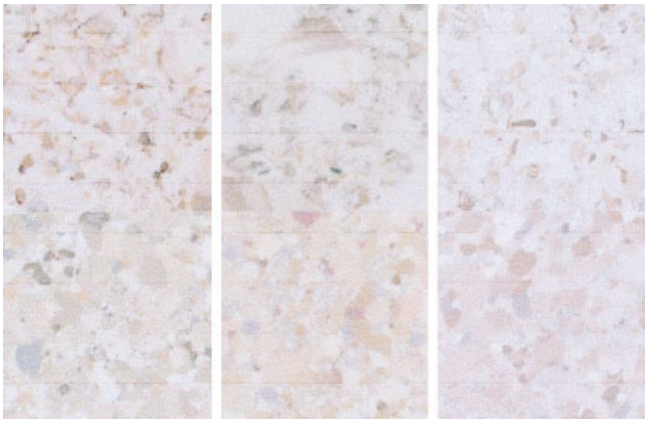

FIG. 4. Vinyl database.

were generated from [1 8 8] and [4 4 4], while the histograms of 512 bins were generated from [8 8 8] and [2 16 16]. Three color spaces (CIELAB, CIELCh, and opponent) were appropriate for investigation in this way, and two quantization models ($1 \times 8 \times 8$ and $2 \times 16 \times 16$) were employed to produce a total of six models. Table I shows the specification of the quantization levels in the three axes of each color space.

Categorical Judgment Experiment

Four databases (containing images of carpet, cork, linoleum and vinyl samples) were selected[19] and six sample images (each of 400×400 pixels) were chosen from each database. These databases were selected for the study because of their abstract nature (the randomized and smooth patterns eliminate human judgments based on the semantic meaning of images) and on their variety of colorimetric and spatial patterns. From each image, 30 smaller images (200×200 pixels) were randomly selected to yield a total of 720 images (4 databases $\times$ 180 images/database), which were

FIG. 5. Observer interface of the category judgment experiment.

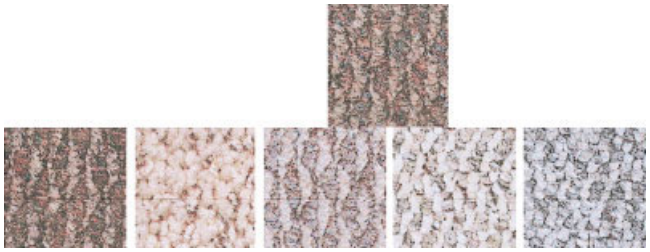

FIG. 6. Session 1: Image set used in the rank order experiment.

then investigated in this study. Figures 1–4 show some typical images that were used.

Six observers participated in the psychophysical experiments of category judgments[20] and each observer made 720 trials. In each trial, two images were randomly selected from the 180 samples of one of the databases and were displayed to the observers, as shown in Fig. 5.

The observer's task was to judge the visual difference between the pair by assigning a category of 1–5, where category 1 represented the most difference and 5 represented no difference. Observers were given no guidance about whether to focus more on the spatial or chromatic differences between the images. Comparisons of the performance between computational models with different combinations of color spaces and quantization levels in fitting these psychophysical results were carried out to quantify their ability to predict human judgment of image similarity of real complex images.

Rank Order Experiment

The purpose of this experiment was to collect psychophysical data on similarity measurement, by ranks of a series of images from different texture images. In this experiment, five target images and twenty comparison images (of 200×200) pixels were selected from the four texture types (containing images of carpet, cork, linoleum and vinyl samples). To investigate the effect of different combinations of the comparison images, some images with particular characteristics were selected and displayed repeatedly in some experiment sessions. Varieties of image set were investigated based upon their color and spatial properties.

The target and comparison images of each session are shown in Figs. 6–10. The image at the top of each figure is

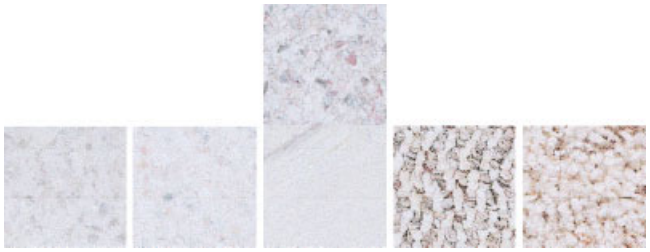

FIG. 7. Session 2: Image set used in the rank order experiment.

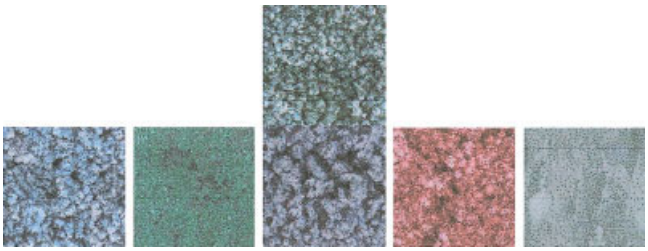

FIG. 8. Session 3: Image set used in the rank order experiment.

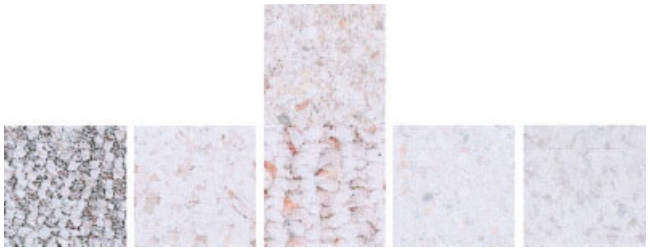

FIG. 10. Session 5: Image set used in the rank order experiment.

always the target image for each session.

Figure 11 shows the initial interface being displayed for observer to make judgment. The target image always appears at the top left corner.

In each session, one target image and five comparison images were shown in each interface and displayed in the fronto-parallel plane of the observers. The observers were asked to rank and arrange a given set of five comparison images, according to their similarities to the target image, in decreasing order. The lower rank represents the most similar. Each observer completed five sessions. Ten observers with normal color vision (confirmed with Ishihara plates) were involved. Totally 10×5 observations were delivered.

The mean rank represents the mean similarity value. The proportion of agreed observer judgment (the number of observers, out of the total of 10, assigned the same rank to the corresponding comparison image) is 74.4%, showing that the observer judgments were reliable and consistent.

The experimental images were selected from different databases with different classes; the ratio of inter- and intraclasses acts as a measure of observer performance. This is computed by the ratio of interclass images to intraclass image being selected by observer as the first rank. Out of 10 observers to make 5 judgments, this ratio is 49:1. In general, images that came from the same class as the target were judged with higher ranks. The high ratio of interclass to intraclass indicates that the observer judgment is reliable.

Model Evaluation of Categorical Judgment Experiment

An estimate of the inter-observer agreement was obtained by computing the standard deviations of the category values over all observers. The frequency distribution of category values will have a high degree of central tendency when observer agreement is good. Confidence intervals were computed from the standard deviations, and interobserver agreement was found to be promising, with 95% confident intervals for the mean category judgment in the range ± 0.336 (cork database) to ± 0.516 (linoleum database).

The average psychophysical response among observers (mean category value) for each image pair was compared with the value of the histogram intersection computed for each of the models. The square of the product-moment correlation coefficient, r^2, was calculated and used to represent the proportion of variation in the psychophysical results (the mean category value) that is explained by the model (see Fig. 12). The values of r^2 range from 0 to 1; the larger the value the better the model fits the data.

To illustrate a typical color-space effect, the histogram intersection values (for the vinyl database) are shown in Fig. 13, with the highest quantization level (8000 bins), using different color spaces.

When performance over all the four image databases is considered, the CIELAB color space generally performs better than the other color spaces in most conditions of different quantization levels. The HSV, CIELCh, and opponent color spaces also perform well. The performance of the RGB color space is not very good whereas the XYZ color space shows the lowest r^2 value in most conditions.

FIG. 9. Session 4: Image set used in the rank order experiment.

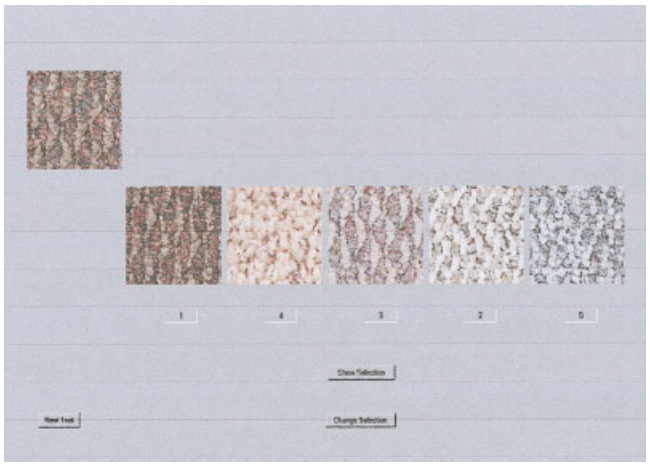

FIG. 11. Observer interface of the rank order experiment when the experiment starts. The target is always on the left top corner. A series of comparison images are on the right bottom.

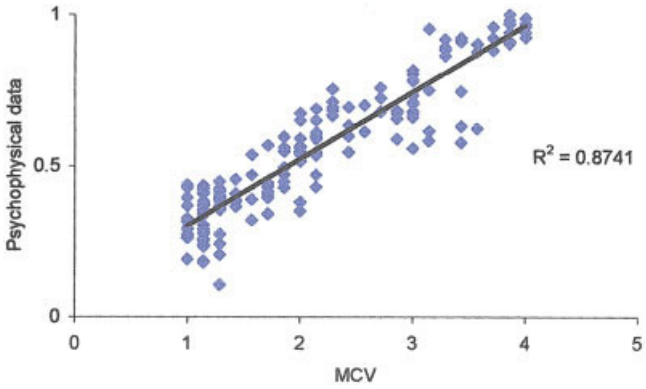

FIG. 12. Example of linear regression representing the relationship between psychophysical data and MCV.

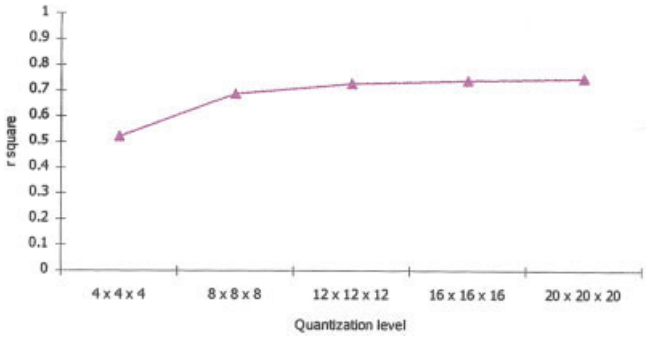

FIG. 14. Linoleum database as an example illustrating quantization of three axes in CIELAB color space with equal number of bins.

The poor performance of the RGB color space may be explained by other color spaces being more perceptually relevant.

Figure 14, using the linoleum database as an example, presents a typical trend of the effect of the quantization level (number of bins) on model performance, for the CIELAB color space. It shows that histograms with more bins give larger r^2 values and perform better in predicting human judgments of image similarity.

As shown in Fig. 14, the similarity performance of the CIELAB space become saturated ($r^2 \sim 0.7$) when the number of bins is increased beyond about 512 ($8 \times 8 \times 8$). Tables II–V show the performance of the models for various quantization levels for the carpet, cork, linoleum, and vinyl databases, respectively. We noted that the model performance saturated at about $r^2 \sim 0.7$ for the carpet and linoleum databases, while only $r^2 \sim 0.5$–0.6 was achieved for the cork and vinyl databases. There is a natural trade-off, however, between performance and computational costs, and we suggest that an $8 \times 8 \times 8$ quantization system with 512 bins may be optimal.

Figure 15 illustrates an example of histogram intersection in different color spaces using the linoleum database for models in which the chromatic channels are given more quantization levels than the achromatic channel. For the case of 64 bins, generally, some color spaces showed an improvement in performance when the available bins were unevenly divided throughout the color channels (see Fig. 15 for an example), but no such effects were evident when the total number of bins was 512.

Model Evaluation of Rank Order Experiment

The estimated similarity responses of observers were correlated with the models by the r^2 values of their linear regression line to show the effectiveness of the computa-

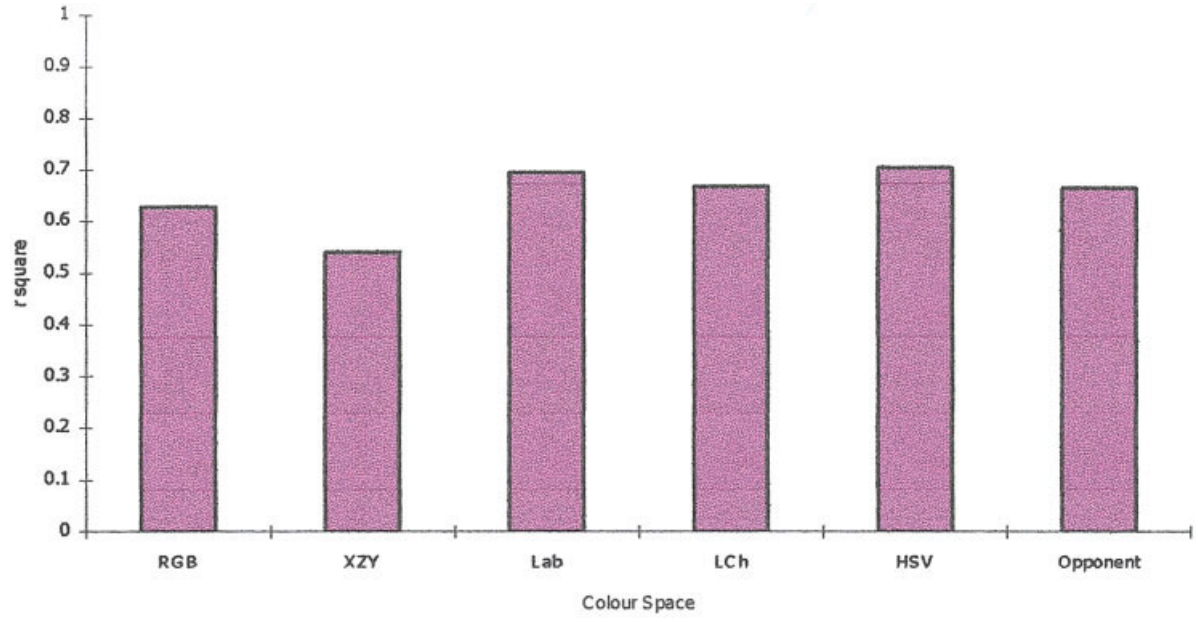

FIG. 13. Vinyl database as an example showing a typical color-space effect to the models performance in predicting human judgment of image similarity using $20 \times 20 \times 20$ quantization levels (80,000 bins).

TABLE II. The r^2 values of the linear regression of the carpet database.

Quantization levels							
Achromatic	Chromatic	RGB	XYZ	Lab	LCh	HSV	Opponent
1	8 × 8	—	—	0.497897	0.485599	—	0.500924
4	4 × 4	0.75482	0.659193	0.624447	0.582684	0.629055	0.764811
2	16 × 16	—	—	0.637283	0.592211	—	0.690083
8	8 × 8	0.730773	0.610778	0.688932	0.632793	0.656832	0.7004
12	12 × 12	0.686272	0.650919	0.666357	0.658434	0.673303	0.682747
16	16 × 16	0.708249	0.628857	0.673703	0.65664	0.634966	0.687563
20	20 × 20	0.645312	0.634749	0.682856	0.672927	0.660825	0.686417

tional models in predicting human judgment, with larger r^2 value representing better model performance.

Generally speaking, the model performance of the rank order experiment agrees with the observation of the categorical judgment experiment that is increasing with the quantization level. With maximum r^2 values of around 0.66, models such as the CIELAB, CIELCh, and HSV models perform better than others, as shown in Fig. 16. It seems that the hue components contribute to such performance. At superficial glance at the experimental images, their hue differences are obvious, this may explain such good performance. The opponent color space does not outperform others, in this model. Again, the XYZ model does not perform well. This can be explained by the perceptual non-uniformity and unintuitive features.

As shown in Fig. 17, the model performances of CIELCh and the opponent color spaces were affected by the quantization; the model with equally-quantized histogram performs better than that with the unevenly-quantized histogram. However, the opposite effect is observed in the model of CIELAB color space, with quantization of [1 8 8] performing better than that of [4 4 4].

CONCLUSION

The purpose of this study was to investigate whether the choice of color space and the number of bins affect the ability of the histogram-intersection technique to make prediction of similarity. The results obtained from comparing the various models with psychophysical data supported the premise that both the choice of color space and the number of bins (quantization level) affect performance.

The psychophysical data of the categorical judgment and rank order experiments were correlated with the computational models. To summarize the model performance of these two experiments, the opponent and CIELAB color spaces generally work better than the other color spaces in most of the quantization levels, choice of database, and even computational models. For the performance of the spatio-chromatic map, the opponent and CIELAB color spaces work well in the categorical judgment experiment, whereas the CIELAB color space performs better in the rank order experiment. The images in the categorical judgment experiment are very spatio-chromatically similar of the same database to evaluate the model efficiency in characterizing

TABLE III. The r^2 values of the linear regression of the cork database.

Quantization levels							
Achromatic	Chromatic	RGB	XYZ	Lab	LCh	HSV	Opponent
1	88 × 8	—	—	0.451647	0.402777	—	0.511958
4	4 × 4	0.46906	0.577046	0.441522	0.449318	0.410246	0.454252
2	16 × 16	—	—	0.545563	0.511991	—	0.490673
8	8 × 8	0.50358	0.544475	0.559567	0.508294	0.465654	0.501472
12	12 × 12	0.518483	0.522851	0.543045	0.522493	0.561618	0.498162
16	16 × 16	0.525476	0.5248	0.570504	0.543508	0.574841	0.505397
20	20 × 20	0.547565	0.514334	0.582783	0.561176	0.586191	0.509253

TABLE IV. The r^2 values of the linear regression of the linoleum database.

Quantization levels							
Achromatic	Chromatic	RGB	XYZ	Lab	LCh	HSV	Opponent
1	8 × 8	—	—	0.626599	0.41233	—	0.630122
4	4 × 4	0.482991	0.586136	0.522639	0.389108	0.557999	0.4251
2	16 × 16	—	—	0.697201	0.45146	—	0.727441
8	8 × 8	0.573356	0.597298	0.688033	0.487747	0.714049	0.667929
12	12 × 12	0.675728	0.558047	0.726639	0.534574	0.742568	0.688044
16	16 × 16	0.536061	0.587168	0.737876	0.54451	0.734261	0.728998
20	20 × 20	0.68745	0.612735	0.746303	0.547908	0.743388	0.747134

TABLE V. The r^2 values of the linear regression of the vinyl database.

Quantization levels							
Achromatic	Chromatic	RGB	XYZ	Lab	LCh	HSV	Opponent
1	8 × 8	—	—	0.587901	0.431615	—	0.458278
4	4 × 4	0.366168	0.342828	0.329828	0.399649	0.321727	0.491232
2	16 × 16	—	—	0.644873	0.525473	—	0.578829
8	8 × 8	0.554004	0.432326	0.635418	0.464175	0.529216	0.580957
12	12 × 12	0.538418	0.482172	0.589152	0.580526	0.64174	0.602479
16	16 × 16	0.617147	0.503857	0.701065	0.594025	0.624101	0.639154
20	20 × 20	0.627067	0.540775	0.695194	0.665365	0.703211	0.661985

image similarity in small differences, whereas the images in the rank order experiment were from different image databases and classes to test the model capability in distinguishing differences within a wider variety of image type. The image dependency may explain the sensitiveness of the model to perform better in some database.

In general, the similarity measurement performance is saturated and optimized when the number of bins reaches 512 (8 × 8 × 8), with a typical performance of $r^2 \sim 0.7$. Further work may be necessary, using different image datasets, to determine whether this result is applicable more widely. We found that the CIELAB color space performs better than the other color spaces that were tested in most conditions (irrespective of quantization level and choice of database), with $r^2 = 0.8$ as its best performance.

In general, the model performances increase with the quantization levels. However, it is worth to consider the natural trade-off. Based on these two experiments, a quantization of [1 8 8] is suggested to be used. There may be some advantage in using relatively more or less bins for the achromatic and chromatic channels.[21] The uniform quantization for perceptually nonuniform color space may be problematic. It is suggested that the weighted quantization may apply to the perceptually nonuniform color space.

On one hand, images of relatively similar color were selected in the category experiment. The model performance of the cork database shows that the color histogram intersection is not so sensitive to distinguish the image differences between images of very similar color distributions. On the other hand, images of very different color, especially hue, were selected in the rank order experiment. The image dependency has been tested. It appears that color-space effect is slightly sensitive to different images, while the quantization effect appears independent on it. For instance, the models with CIELCh and HSV color spaces perform better and demonstrate their sensitiveness in the images of the rank order experiment.

The correlations between model predictions and psychophysical judgements of image similarity were generally in the range of $r^2 = 0.6-0.8$. This level of performance may be sufficient for certain image-indexing applications. However, it is likely that better performance would be possible

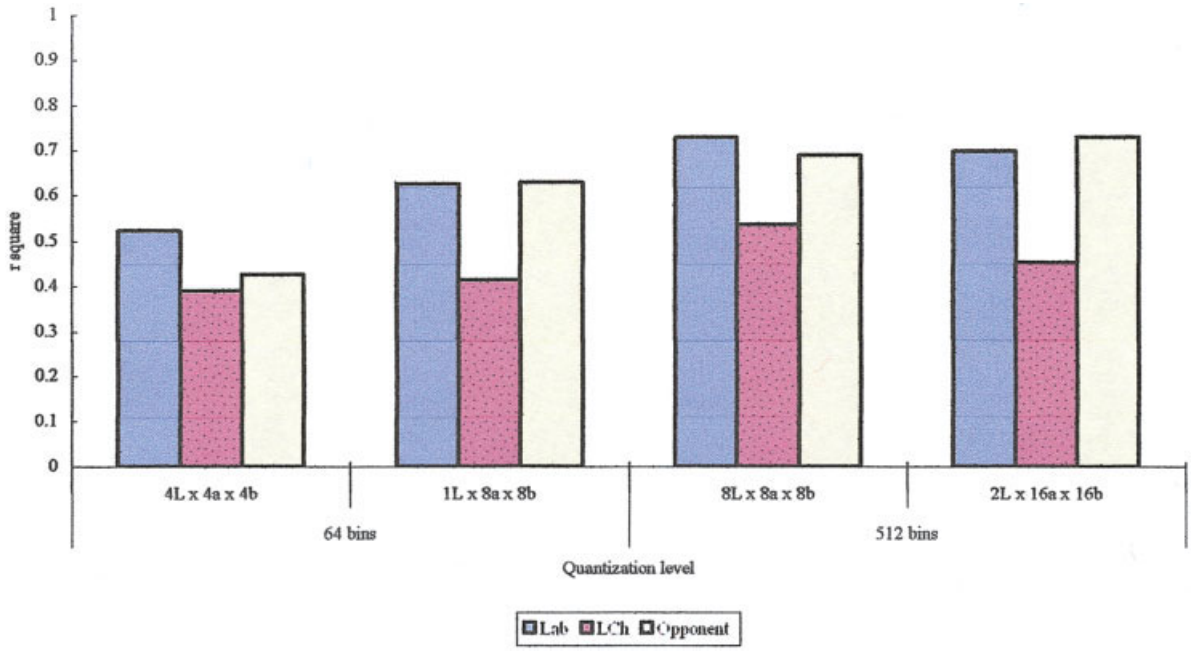

FIG. 15. Linoleum database as an example illustrating quantization of three axes in a different color space with different number of bins.

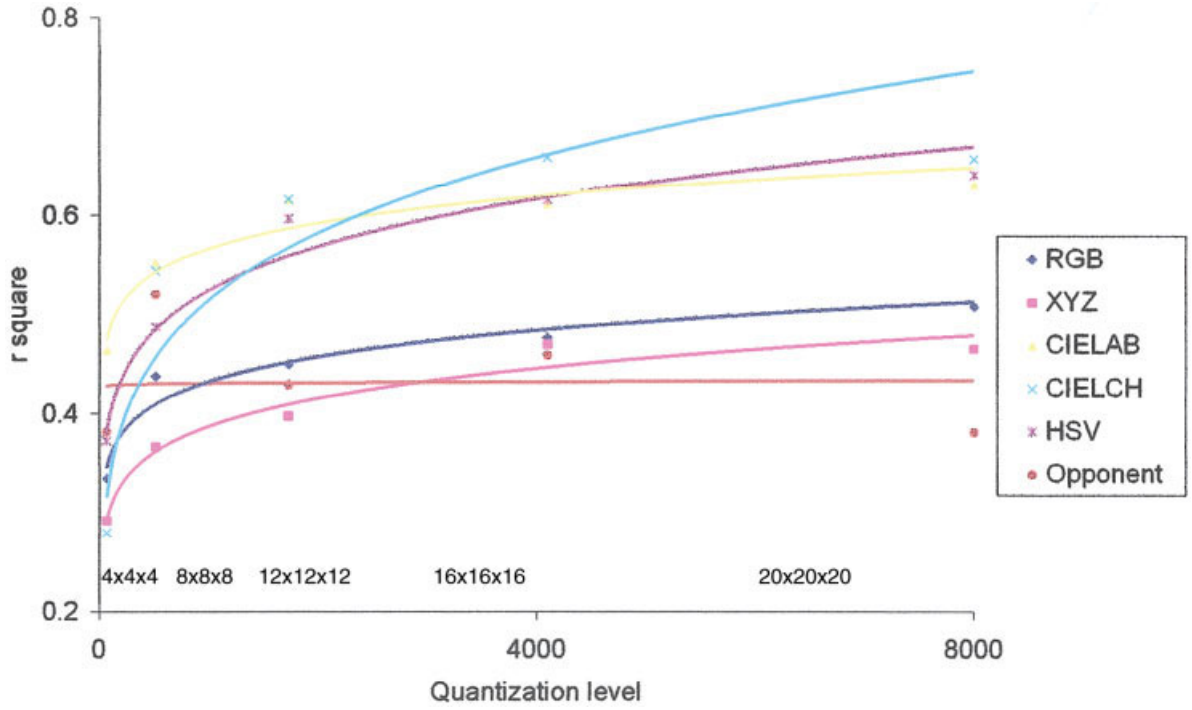

FIG. 16. Agreement of model with psychophysical data for different color spaces and for different levels of quantization.

using models that take into account the spatial properties of the images in addition to the color properties.

APPENDIX: MONITOR CHARACTERIZATION DETAILS

RGB values in a particular set of primaries can be transformed to and from CIE XYZ by three-by-three matrix transform. To transform from RGB to CIE XYZ the digital input RGB values (D_r, D_b, and D_g) must be normalized and linearised according to the GOG model. A typical characterization method was employed, which resulted in the following transform:

$$R = (1.185002Dr/255 - 0.185002)^{2.244708};$$
$$G = (1.187830Dg/255 - 0.187830)^{2.248395};$$

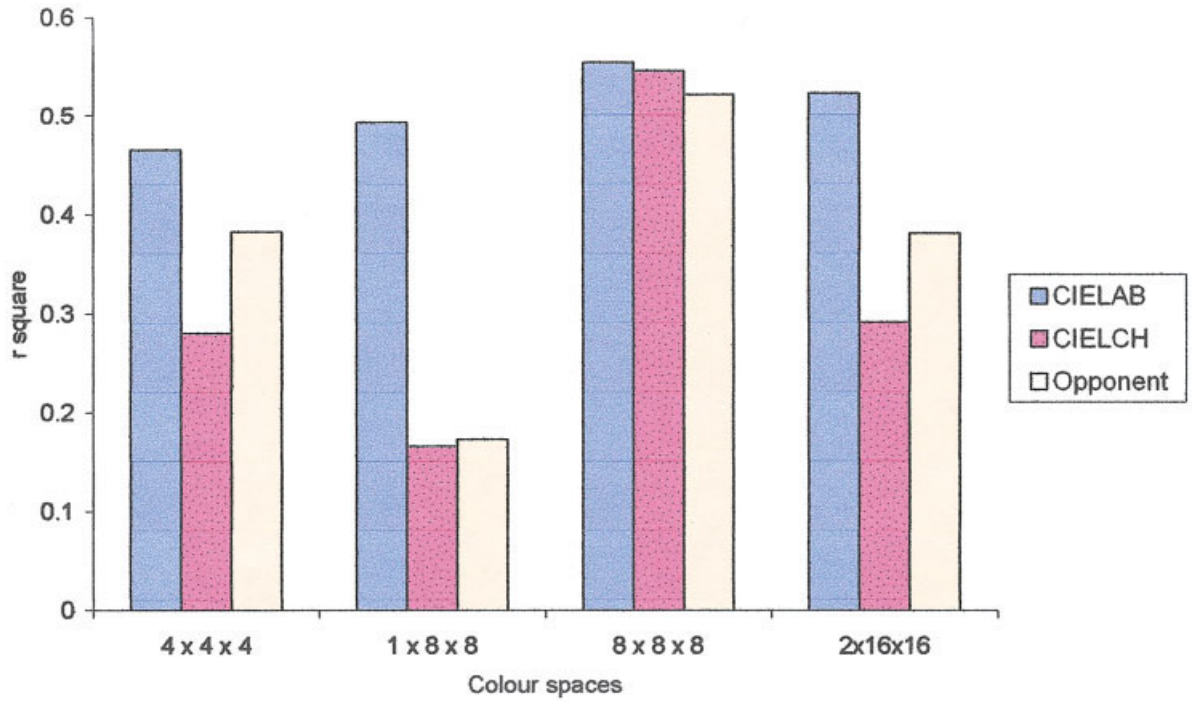

FIG. 17. Quantization effect on color histogram intersection of the rank order experiment.

$$B = (1.229194Db/255 - 0.229194)^{2.282509};$$

The linear RGB values (R,G,B) were then converted into XYZ values using the following linear transforms:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 42.337724 & 33.826257 & 18.834455 \\ 23.755334 & 64.651494 & 12.062589 \\ 2.061681 & 9.472647 & 98.460478 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$X = (X + 0.373846)*70.8/100;$$
$$Y = (Y + 0.384068)*70.8/100;$$
$$Z = (Z + 0.306871)*70.8/100;$$

1. Caelli T, Reye D. On the classification of image regions by color, texture and shape. Pattern Recogn 1993;26:461–470.
2. Wuerger SM, Morgan M, Westland S, Owens HC. The spatio-chromatic sensitivity of the human visual system. Physiol Meas 2000;21: 505–513.
3. Niblack W, Barber R, Equitz W, Flickner MD, Glasman E, Petkovic D, Yanker P. The QBIC project: querying images by content using color, texture and shape. In SPIE Conf. on Storage and Retrieval for Image and Video Databases, 1993, 1908. 173–187.
4. Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R. The virage image search engine: an open framework for image management. In SPIE Conf. on Storage and Retrieval for Image and Video Databases, 1996, 2670. 76–87.
5. Finlayson GD, Chatterjee SS, Funt BV. Colour-texture indexing. Intelligent Image Databases, IEE Colloquium, Dec. 1–6, 1996.
6. Swain MJ, Ballard DM. Colour indexing. Int J Comput Vis 1991;7: 11–32.
7. Park D-S, Park J-S, Kim T, Han JH. Image indexing using weighted color histogram. ICIAP. 1999, p 909–914.
8. Mehtre BM, Kankanhalli MS, Desai Narasimhalu A, Man GC. Colour matching for image retrieval. Pattern Recogn Lett 1995;16:325–331.
9. Wan X, Kuo CCJ. Colour distribution analysis and quantization for image retrieval. SPIE Proceedings, Vol. 2670. 1996.
10. Wang JZ. Integrated region-based image retrieval. Boston: Kluwer Academic Publishers; 2001.
11. Pratt WK. Digital image processing, 3rd edition. New York: John Wiley and Sons; 2001.
12. Sproson WN. Colour science in television and display systems. Bristol, UK: Adam Hilger; 1983.
13. Feiner SK, Hughes JF. Computer graphics: principles and practice. Reading, MA: Addison-Wesley; 1996.
14. Berns RS, Motta MJ, Gorzynski ME. CRT colorimetry. Part I: Theory and practice. Color Res Appl 1993;18:299–314.
15. Berns RS, Motta MJ, Gorzynski ME. CRT colorimetry. Part II: Metrology. Color Res Appl 1993;18:299–314.
16. Berns RS. Methods for characterizing CRT displays. Displays 1996; 16:173–181.
17. McDonald R. Colour physics for industry. West Yorkshire, England: Society of Dyers and Colourists on behalf of the Dyers' Company Publications Trust; 1997. p 138.
18. Smith VC, Pokorny J. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. Vision Res 1975;15:161–171.
19. http://www.ifloor.com (Last accessed 5 February 2004).
20. Gescheider GA. Psychophysics: the fundamentals. Mahwah, NJ: Lawrence Erlbaum Associates; 1997.
21. Berens J. Image indexing using compressed color histograms. Ph.D. Thesis, University of East Anglia, Norwich, 2003.